

Using visual information analysis to explore complex patterns in the activity of designers

Philip Cash¹, Tino Stanković², Mario Štorga³

Affiliations:

¹ Technical University of Denmark, Denmark

² ETH Zurich, Switzerland

³ University of Zagreb, Croatia

Please cite this article as: Cash, P., et al., Using visual information analysis to explore complex patterns in the activity of designers, *Design Studies* 35 (1) (2014) 1-28

<http://dx.doi.org/10.1016/j.destud.2013.06.001>

Technical University of Denmark
Produktionstorvet
Building 424, room 122
2800 Kgs. Lyngby
Denmark

Email: pcas@dtu.dk
Tel: (+45) 45255563
Fax: NA

Using visual information analysis to explore complex patterns in the activity of designers

Abstract

The analysis of complex interlinked datasets poses a significant problem for design researchers. This is addressed by proposing an information visualisation method for analysing patterns of design activity, qualitatively and quantitatively, with respect to time. This method visualises the temporality of interrelations between interlinked variables and, as such, can be applied to a range of datasets. By providing a statistical analysis of the networks' growth the proposed method allows for the modelling of complex patterns of activity. Throughout, the method is demonstrated with respect to a fully realised example of information seeking activity. The core contribution of the proposed method is in supporting the analysis of activity with respect to both macro and micro level temporal interactions between variables.

Keywords: Research methods, design research, design science, case study, experiment, visualisation analysis

This paper demonstrates a method for using information visualisation and statistical analysis to explore complex patterns in the activity of design practitioners, over time. Within design research there has been a major and long lasting research focus on the activities (Dorst & Dijkhuis, 1995; Pedgley, 2007) and behaviours (Coley, et al., 2007; Hyldegard, 2009) of the design practitioner. This has been supported by a range of empirical studies (Goodman-Deane, et al., 2010; Robinson, 2010) based on both quantitative (mainly realist or positivist) and qualitative (mainly constructivist) approaches (Lethbridge, et al., 2005).

Despite this research focus, much of the complexity of designer activity remains inscrutable due to the limitations of current quantitative approaches. This is particularly problematic in the context of the evolution of activity over time. Current approaches typically use protocol based techniques (Chai & Xiao, 2011; Gero & Mc Neill, 1998) which, Gero and Mc Neill state can be generally characterised as developed during the analysis. However, Gero and Mc Neill (1998) go on to highlight that in design this type of analysis is typically based on a segmented timeline where protocols are split up over time based on various factors, such as information activity (Robinson, 2010) or artefact use (Huet, et al., 2007). The idea of a segmented timeline is consistent with all of the major approaches to this type of analysis in the design research literature. For example, Goldschmidt and Tasta (2005) discuss the foundational work of Wang and Habraken (segmented timeline) to help contextualise design moves – steps that parse the design process into short verbalisations on a timeline. Further, Gero and Mc Neill's (1998) 'episodes' again form a

decomposition over time which can themselves be segmented based on, for example, domain, micro strategy (Gero & Mc Neill, 1998) and function-behaviour-structure (Kan, et al., 2011). However, using this type of analysis does not readily allow for the quantitative assessment of patterns of activity or the evolving relationships between activities over time. A more detailed example of this type of issue can be found when we consider the work of Dong et al. (2012), whose analysis of design cognition encountered a number of interlinked variables including *team*, *architect*, *culture*, *energy* and *health*. In this case, Dong et al. chose to analyse this data using several measures including the rate and duration of each variable, a graphical representation of the subject's communication acts and a numerical analysis of the total percentages of communications and moves with respect to Dong et al.'s concept of frames. Nevertheless, despite the clear and useful contribution of Dong et al.'s work there are difficulties in linking multiple variables and identifying patterns beyond those at the individual variable level. Thus, a key limitation of current work is the analysis of multiple variables in a complex setting.

Therefore, by focusing on the common feature of this type of analysis – decomposition of the design process into distinct periods in time – the method proposed in this paper builds generalisability across research foci and analysis approaches by giving generic steps for the visual analysis of any activity based dataset decomposed over time. In this sense the proposed method is both complementary to and expands upon the ideas expressed in the development of Linkography (Goldschmidt, 1992; Kan & Gero, 2008). In this context visual information analysis offers the potential for examining protocol data in a new way to allow the quantitative analysis of these patterns and relationships (Nguyen, et al., 2012). Despite this possibility, such an approach has not been developed for characterising activity in the context of design research.

Based on these considerations this paper proposes a method for using visual information analysis to quantitatively characterise designer activity. In this context a key contribution of the proposed method is in elucidating the temporal growth of patterns of activity – a key deficit in existing approaches. The paper first describes relevant background elements (Section 1) before outlining the overall network creation approach (Section 2). Next the study used to exemplify the approach is outlined – using information seeking as an example – (Section 3), and the proposed approach described fully (Section 4) before the results are discussed in detail (Section 5).

1 Background

Two areas are examined in this section. Firstly, it is necessary to understand the current state of the art of visual information analysis in design and, in particular, its use in design research (Section 1.1). Secondly, in order to provide a meaningful demonstration of the approach, the relevance of information seeking activity and its associated research is discussed in detail (Section 1.2).

1.1 Visual Information Analysis

Visual information analysis aims to enable the interpretation of complex datasets, which would otherwise be inscrutable. To this end, the goal of information visualisation can be defined as: *“to explore data in order to gain understanding of the data and the phenomena behind”* (Purchase, et al., 2008)(p.54). Further, cognitive psychology has established that, for any given task, it is easier to process complex structures visually rather than relying on working memory alone (Ware, 2004). The natural ability to deal with complex digital repositories can be augmented by suitable visualisation techniques, applied to support interpretation through recognition of patterns in the images. Thus information visualisation offers the potential to elucidate salient structures, patterns or relationships, which would otherwise not be accessible using traditional presentational graphics or information communication techniques. Information visualisation tools can display information as spatially organized and interactive, alleviating much of the effort required in building mental models of phenomena.

In general, data and information visualisation techniques complement statistical analysis to facilitate understanding of complex sets of data (Unwin, et al., 2006). In this context, visualisation is primarily used to represent and qualitatively explore large sets of data, simulations or analysis results. For instance, a value-path representation of multi-objective optimisation results enables an analyst to qualitatively assess the effectiveness of the optimisation algorithm and to establish a trade-off among solutions. This qualitative approach has been extended to include interactive elements through which the analyst can explore complex relationships. For example, Parmee and Abraham (2004) use a cluster-oriented genetic algorithm to provide an interactive user-centric visualisation environment in which the designer can interact with the system directly.

Another key role of visualisation in the design process is in facilitating project management. This typically takes the form of table-based output summaries suitable for personal use and of limited scope in terms of complexity. Even in relatively simple systems the accumulation of information in

this format rapidly becomes unwieldy and difficult to systematically search. For example, Salustri et al. (2008) highlight this issue in their examination of computer-based diagramming tools. Finally, it is important to note that using visualisations for analysing data is not new (Waddell, 2002) but there remains a substantial challenge in adapting and applying them in a given context – something that presents a significant opportunity in the context of the design domain.

Despite this varied body of work some significant issues still remain in representing complex and evolving information structures – primarily the combination of both qualitative and quantitative elements. A new visualisation tool – OrganicViz (www.organicviz.org) – addressing this problem was introduced by Stanković et al. (2012) to visually support traceability in the evolution of engineering information. Here, an organic information visualisation (Fry, 2000) is employed to reveal the dynamics of engineering information use as a complex heterogeneous network of traced information elements and objects interconnected by semantic links (Pavkovic, et al., 2013; Storga, et al., 2011). This dynamically evolving network, supported by graph layout algorithms, offers a narrative explanatory approach for the examination of information structure evolution and interaction. This is true even for extremely complex systems and also allows for the quantitative analysis of the resulting network.

Further to the work focused on developing visualisations for the design process there have been several attempts to use this type of approach to support the scientific process in design research. For example, Vande Moere et al. (2008) highlight the unsuitability of conventional organisation charts and graphs for providing understandable representations of team performance at a social level when studying collaboration. In a more recent example, Nguyen et al. (2012) used techniques from bioinformatics to uncover Computer Aided Design (CAD) best practice and quantitatively describe the processes in which CAD systems are used. As such, it is possible to see the potential of using visualisation approaches for decomposing complex systems in the design research context. In particular, this type of approach could offer significant advantages in the examination of designer activity or behaviour.

Building on this premise the OrganicViz tool offers advantages over existing approaches in the research context due to its ability to quantitatively explore complex networks of linked activities such as those found in the activity of design practitioners during a design episode. For graph layout the OrganicViz tool employs a Barnes-Hut algorithm (Barnes & Hut, 1986) to optimize

calculation time, providing a network that can be interacted with in real time. Further, the tool employs a fast hierarchical agglomeration algorithm for community identification (Clauset, et al., 2004), which was modified to accommodate directed multigraphs. Additionally, to facilitate analysis the OrganicViz tool employs real-time network properties calculation including mean degree, regular and harmonic geodesic distances, betweenness centrality, various taxonomy/ontology dynamical filtering options, nodal neighbourhood exploration, labels to provide semantics, etc. The OrganicViz tool has been used to support visually augmented analysis of various feature of design research, most notably: the evolution of content in the design research field (Stankovic, et al., 2012), after-sales services for supporting open innovation in a product/service-system (PSS) life cycle (Storga, et al., 2013), and the evolution of the structure of organisational knowledge in a non-governmental organisation based on a recursive email interaction analysis. This work extends the application of OrganicViz to the analysis of the engineering design processes, in particular using the example of information seeking activity.

As such, this work builds on the OrganicViz tool, utilizing dynamic and organic complex network visualisation, to develop an approach for facilitating the quantitative and qualitative analysis of complex activity patterns. However, in order to effectively demonstrate the approach, a suitable subject is required. In this context the area of information seeking research has two main advantages. Firstly, the multiple interactions between the designer and the various information sources (Robinson, 2010) provide an accessible subject in which complex networks of activity are likely to arise. Secondly, information seeking is a key area in design research that has, to date, not been described at the level of quantitative detail offered by this approach. Both of these aspects are explored in more detail in the following section.

1.2 Information Seeking

Information seeking plays a key role in engineering design practice (Reed, et al., 2011; Robinson, 2010) accounting for between approximately 20% (Court, et al., 1998; Reed, et al., 2011) and over 60% (King, et al., 1994; Robinson, 2010) of engineers' time. Although this has been the case for many years (Cave & Noble, 1986; Puttre, 1991) the increasing importance and impact of Internet-based information seeking is, as yet, little understood. For example, Robinson (2010) classes the Internet as a single information source.

Two key studies have highlighted the importance of the Internet as one of the primary sources of information for engineering designers. Allard et al. (2009) identify Internet based information as the second most prevalent ‘information activity’, only being superseded by the engineers’ own working documents. This is further supported by Kwasitsu (2004) who found that the Internet was considered the fourth most important source after ‘*people in own business group*’, ‘*personal files*’ and ‘*personal memory*’.

As a result of these and other such studies there have been several attempts to understand the interaction between the user and source. For example, Holscher and Strube (2000) emphasize the multifaceted interplay between browsing and search engine activities. Further to this, Keller et al. (2007) highlight the complexity of Internet based information seeking activities. However, due to this complexity it has been difficult to explore this activity in detail, with even the most recent studies focusing on characterising activity in terms of total time allocation (aggregated over a whole study) (Reed, et al., 2011).

The limitations of these current quantitative approaches have restricted the scope of investigation with respect to elucidating the complexity of information seeking activity. For example, Kwasitsu (2004) found overall differences in seeking activity related to job role but was unable to further explore the nature of these differences. As such, a new approach is needed to decompose the complexity of information seeking (and complex designer activity in general) in order to more effectively support the evolving needs of engineering designers and design researchers. In particular, as interactions and activity become less person focused and increasingly involve complex technological systems and resources such as the Internet.

To address this, the proposed visualisation approach was created, and is outlined in the next section before the study used in its demonstration is described in Section 3.

2 Network Creation

The focus of the proposed approach is the use of visualisation techniques to explore the relationships between activities, both qualitatively and quantitatively. In order to demonstrate this, a typical protocol analysis dataset was generated based on the study described in Section 3. Although this example is used throughout this paper for illustration purposes, the proposed

approach is generic and can, therefore, be applied to any such dataset (see Step 1). As such, the visualisation and analysis comprises three main steps. These steps are summarised in Figure 1.

- 1 Select an appropriate dataset for analysis. The approach is limited to situations where data can be defined taxonomically and can also be related to each other. As such, the taxonomy of variables and their relationships are defined based on the specific case. The most basic case is that each term has duration and, as such, can be related in time. In this case a protocol type analysis was used and relationships were defined as overlaps in time.
- 2 Network creation. This step involves mathematical modelling to translate the recorded data types considered in the protocol analysis into the network.
- 3 The network structure's interactive visualisation provides qualitative representations of the network structure dynamics to enable explorative information analysis of the recorded information seeking process. To complement qualitative analysis a rate of growth analysis of the network structure with respect to node and edge addition/rewiring is performed and represented quantitatively with regular chart plots.

Step one is typical of existing approaches, being core to protocol analysis, and is therefore not discussed further here (Gero & Mc Neill, 1998). Step two establishes the foundation for further analysis by relating the defined codes together in a mathematical model as detailed in Section 4. Step three describes the visualisation and analysis of both the qualitative and quantitative data. As such, Sections 4.1 and 4.2 respectively deal with the qualitative and quantitative elements of the visualisation, detailing the approach and illustrating each section with example results from the studies.

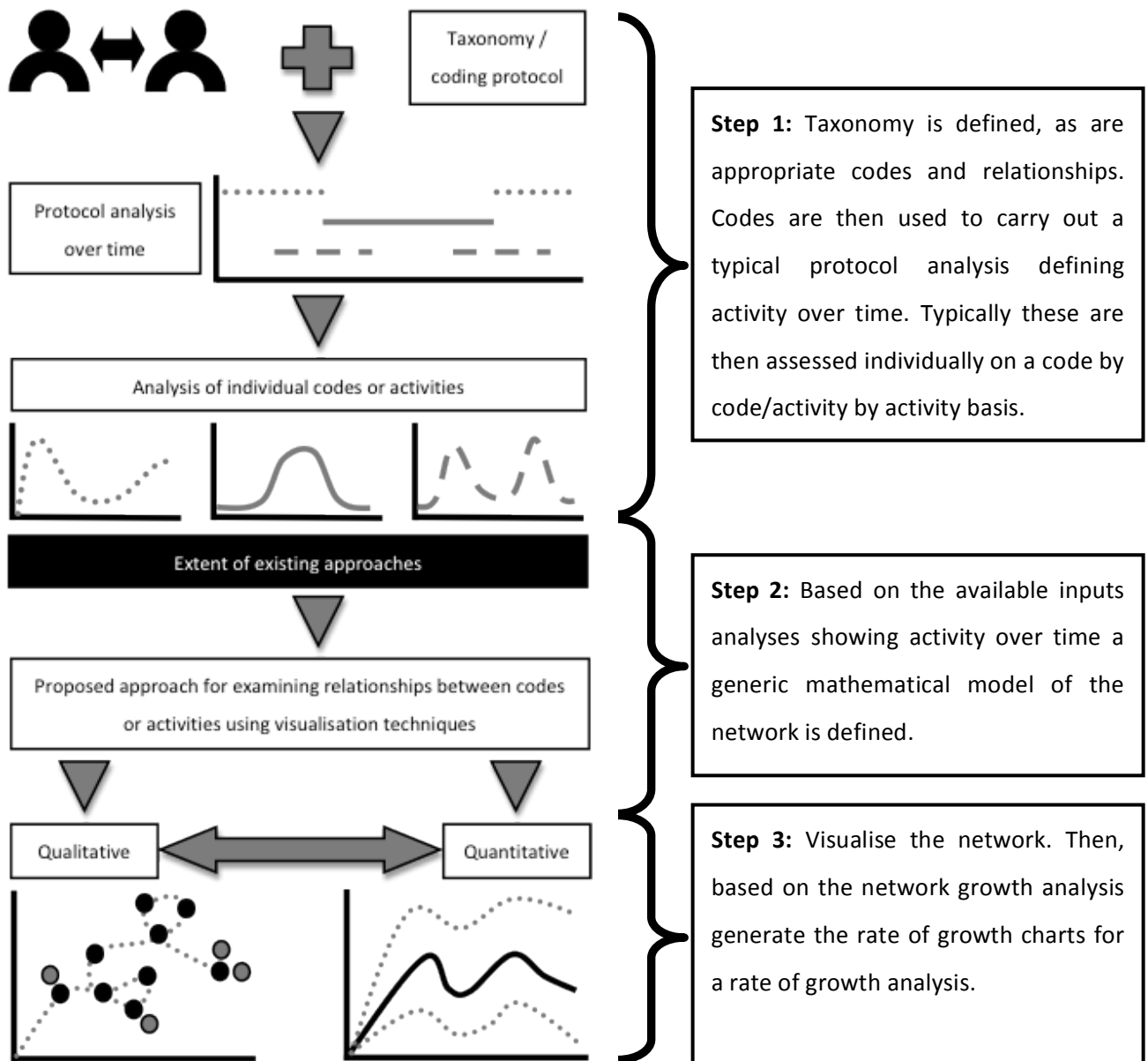


Figure 1: Overall visualisation and analysis steps and its relation to existing approaches

3 Research Design

This section deals with three main elements. Firstly, the two studies are outlined (the observational study in Section 3.1 and the subsequent experimental study in Section 3.2). Secondly, the core coding schema is introduced in Section 3.3.

3.1 Observational Study

The observational study followed the work of an engineering design practitioner in a UK-based Small to Medium size Enterprise (SME) for one week. A UK SME was selected at this stage for two reasons – they are the typical focus for design research and they make up the majority of UK based companies, accounting for 99.9% of all enterprises and 58.8% of private sector employment (White, 2011). The participant was selected using two steps. First volunteers were recruited,

producing a subset of five (out of seven) engineers. Secondly, one of the volunteers was randomly selected for the study in order to avoid further selection bias (Torgerson & Torgerson, 2003). Although a fully randomised selection regime offers the best approach (Torgerson & Torgerson, 2003) voluntary screening was necessary due to the level of observation undertaken and was, therefore, the best pragmatic option available.

The observation approach recorded the participants' computer screen (using the Panopto software (Panopto, 2011)), video footage of their workstation and their logbook activity (using a LiveScribe pen (LiveScribe, 2011)). This was implemented over a four-week period with the final week being used for the study, allowing the participants to acclimatise to the setup (Podsakoff, et al., 2003). Based on this data, a period of comparison was identified, defined as the longest continuous period of information seeking. This was then used as the basis for developing the experimental study and the codes for the *general sources* (Section 3.3). The period identified constituted 100 minutes of Internet based information seeking, focusing on feasibility stage information with respect to a small electro-mechanical product in preparation for later brainstorming and design review tasks.

3.2 Experimental Study

The experimental study used two groups of participants – students and practitioners. In this case twelve students (from 40) and three practitioners (from seven – also drawn from the SME) were randomly selected for the study. Students were selected from a final year product design and development course. This ensured that each participant's experience and background were relatively homogeneous, with an average of 10 months industrial experience and 4 years academic training.

The experimental study was based on the period described during the observational study, *i.e.* seeking information to support the design of a small electro-mechanical product to be used in subsequent brainstorming and design review tasks. The study was carried out as follows:

Background questionnaire >> Briefing >> Free information seeking (50 min) >> Debrief

The brief given for the experiment left the participants unconstrained with regard to feasibility, cost and scope – similar to the feasibility stage of product development (based on the observation study) – and was as follows:

"You are to design a universal camera mount for use on an aerial vehicle. The aerial vehicle is to be used by an amateur photographer, primarily to take still photos. Using any means available to you, search for and note down information that may help."

The participants were also instructed that the purpose of the searching activity was to support a later brainstorming task to be carried out as part of a larger study. The specific task was selected because it provided an analogous task to the one recorded during the observational study and neither the students nor practitioners were familiar with the brief. Participants were provided with a fully equipped library of physical catalogues and other design literature, a computer workstation and logbook. As in the observational study the computer screen, workstation and logbook were recorded. Participants were tested individually, not being permitted to talk to the other participants until the study was complete.

3.3 Coding

With the data collection complete, the three sources (screen, workstation video and logbook recording) were converted to .mp4 files and synchronised using the VCode (2011) coding and analysis software (Hagedorn, et al., 2008).

Based on the assessment of current information seeking literature in the engineering design domain (Section 1.2) it becomes apparent that Internet based information seeking requires further elaboration before effective coding can take place. This is supported by the wealth of research examining the use of the Internet as an information source in other contexts, such as, healthcare information (Berland, et al., 2001; Jadad & Gagliardi, 1998) and general search activity (Blandford & Attfield, 2010; Peterson & Merino, 2003). Based on this work in other domains two key points emerge.

Firstly, at the activity level there are a number of information activities composing the information seeking/usage process. These can be related to what Belkin et al. (1982) term the information journey (recognizing need, finding information, interpreting/validating information and using that interpretation). Elaborating these for the engineering design domain based on the work of Wasiak et al. (2010) and others results in the following: Recognizing is a prerequisite for the information seeking activity itself and, therefore, is excluded from this study (although it could be incorporated in a broader investigation). Finding information is characterised as both seeking and requesting

(based on Aurisicchio et al. (2010)) while seeking is decomposed into finding source and finding within source based on Robinson (2010). Finally, Borlund (2003) highlights that when analysing information seeking activity it is advantageous to associate sources with the underlying need rather than the query. As such, interpreting and using interpretation have been decomposed into solving, interpreting and evaluating (based on Wasiak et al. (2010)) in order to associate the sources with their specific use and therefore the underlying need.

Secondly, there is no one taxonomy accepted for the classification of internet based sources, with Blandford and Attfield (2010) highlighting Grounded Theory (Glaser & Strauss, 1967) as a key approach in this context. Thus, at the source level, an initial list of core sources was compiled in order to guide a grounded assessment of sources accessed during the observational study. This was mainly based on three works (Hertzum & Pejtersen, 2000; Oh, et al., 2009; Peterson & Merino, 2003) and resulted in the following initial terms:

- Search engines
- Organizations' websites
- Wikipedia (and other wikis)
- Commercial (catalogues and specific commercial articles)
- Experiential sites (product trials, user communities, real usage information)
- News and media
- Personal websites and social media (including experts and friends or other social circles)

These formed the basis for the assessment of the observational study data and were subsequently refined and synthesized to form the key codes for the engineering design context to be used in this study. Table 1 summarises the finalised codes used in the analysis and subsequent visualisation (Section 4). Finally, further detail was provided by defining specific sources at the page level e.g. *Amazon.com/catalogue search: tripod/page 1* would count as a different specific source from *.../page 2*.

Table 1: Information seeking codes and descriptions

Group	Code	Description
Activity	Seeking/requesting	Finding information / direct requests to another party to provide information
	Solving	Involves searching, gathering, creating, developing solutions
	Evaluating	Judging the quality, value and importance of something
	Interpreting	Assigning meaning or value to information
	Find source	Searching for information relating to where specific product information is available

General sources	Find within source	Searching within a specific website for information related to the product
	Search engine	A website that retrieves data, files or documents from the whole internet
	Catalogue	A website that provides a list of items, specifically for sale – entries can also include technical information
	Tech (technology) article/blog	A website giving general commentary on products, technologies and other technical literature in an informal manner
	Supplier article	A website giving commentary on products or technologies written and hosted by the supplier of said product/technology etc.
	Forums	A website hosting a message board
	Expert/supplier	A specific acknowledged expert or product supplier
	Social media	A website hosting user uploaded and accessible content for the purposes of social interaction
	Wiki	A website developed collectively which allows users to add and edit content but with a specific focus such as informing
	Patent	A website displaying a specific patent document
	Standard	A website displaying a specific standard such as the British standards
	Specific source	A specific webpage

Once coding was complete the data was exported as a timeline described in terms of activity/source, start time and duration. As such, once synchronisation was complete, coding was undertaken as follows:

- 1 Overall activity was broken down in order to identify periods of information seeking activity, information requests, direct information use and other unrelated activity.
- 2 Information seeking was decomposed by general Internet sources (although other physical resources were offered these were not used and as such are not included as codes).
- 3 Each source was then described at the webpage level to identify when specific sources were used or reused.

4 Approach to Visualisation

As outlined in the three steps in Section 2, once an appropriate data set was established (Section 3), the next step is to develop the mathematical model and then to generate representations of the structure to allow further analysis. As such, this section first introduces the overall mathematical model used before exploring the two types of representations in Sections 4.1 and 4.2.

Based on the data types considered in the protocol analysis (see Table 1) the network G will be heterogeneous, i.e. it may involve more than one vertex or edge type. Further, multiple edges between activities and sources can occur, depending on the frequency with which an information source is accessed. A labelled multigraph was used as the basis for establishing the network in

order to meet the required level of modelling expressiveness. In the context of the example study, the generated network is considered as an edge and node dynamic network (Harary & Gupta, 1997). It represents the information seeking activity – linking sources to activities in time instances i of the recording session. In order to generate the network a number of definition sub-steps are required. Each step is illustrated using the information seeking example outlined in Section 3, which are boxed for clarity.

Step 1. Define node types to be applied for creating the network G . This can be based on theory or generated based on a grounded assessment of the data as in this example.

Let the codes, defined in the taxonomy, constitute a set of graph node labels \sum_V . In this case there are two subsets (types) – activities \sum_A and information sources \sum_S , thus $\sum_V = \sum_A \cup \sum_S$. Each of the subsets is defined as follows:

- Activities A from set \sum_A that are carried out: seeking/requesting, solving, evaluation, interpreting, finding source and finding within source.
- Information sources S from set \sum_S that are accessed: search engine*, catalogue*, tech article/blog*, supplier article*, forums*, expert/supplier*, wiki*, patent*, standard* and social media*. In order to distinguish among different sources of the same type i.e. the specific sources, the asterisk is replaced by an ordinal number when labelling the network nodes.

Step 2. Define edge types to be applied for creating the network G . Edges can again be based on theory or other relationships without specific a priori definition as in this example.

Let \sum_R be a finite set of graph labels. These are introduced to provide relationship semantics for the network. In this case there are two graph edge labels in \sum_R :

- *Temporal* to show time context, i.e. precedence order of the activities.
- *Mapping* which denotes a direct link, i.e. accessing the individual information source.

Steps 1 and 2 involve two additional mappings: from node labels \sum_V to each node v in node set V and from edge label in \sum_R towards each edge e in edge set E . When these mappings are established the graph nodes v from node set V will be referred to as the taxonomy of coded activities or sources while the edges e belonging to the edge set E will be referred to as the relationships or links.

Step 3. Create a network G based on: the available label sets, their mappings to nodes and edges, and the set of conditions required to define nodes and edges. Let a rule set ρ cover the conditions required to define nodes and edges and all of the specificities that might occur with respect to the input list (protocol data) required to establish an edge. The rule set ρ involves using a range of mathematical approaches depending on research focus.

In this case the recorded design session was broken down into lists of activities and information sources with their session entry point and duration defined in integer time units using the *VCode* (2011) software. Thus, for each activity A in \sum_V and information source S in \sum_S the session entry point map is defined as $t_0 : \sum_V \rightarrow \mathbb{N}$ and the duration time map as $t_d : \sum_V \rightarrow \mathbb{N}$. Based on this input list and the established maps, it is possible to construct an *information seeking network* for any time point during the session obeying the following rules from the rule set ρ :

- Activities are not considered as unique, thus, whenever they appear in the input list, a new node – labelled according to \sum_A – is added to the network.
- Information sources are considered as unique (necessary for identifying iterations where a specific source is revisited one or more times). As such, there is only one labelled node in the network for each label in \sum_S .
- Activities A_1 and A_2 are connected by a *temporal edge* e_t , i.e. $e_t = \{A_1, A_2\}$ if the two activities overlap in time. Given the entry points $t_0(A_1)$, $t_0(A_2)$ and durations $t_d(A_1)$, $t_d(A_2)$ and assuming that the nodes entry points create a well ordered set, activity A_1 is in time overlap with A_2 if the following holds: $t_0(A_2) \in [t_0(A_1), t_0(A_1) + t_d(A_1)]$
- Activity A and information source S are connected by a *mapping edge* e_m , i.e. $e_m = \{A, S\}$ if they are in time overlap satisfying $t_0(S) \in [t_0(A), t_0(A) + t_d(A)] \vee t_0(A) \in [t_0(S), t_0(S) + t_d(S)]$. As a consequence multiple edges between an activity and source may appear if the source was revisited within the duration of the activity.

4.1 Network Visualisation

The OrganicViz tool (Stankovic, et al., 2012) offers a graph-based information visualisation that allows information to be displayed spatially and organized semantically. It lets the analyst browse interactively, navigating to uncover information, and to establish emergent patterns as network motifs. To obtain a qualitative and rich description, the tool utilizes organic information visualisation (Fry, 2000). This uses the metaphor of living (organic) systems to establish a visualisation technique by mimicking organic properties and behaviour. Since organic systems are, by definition, open and information processing systems (Mitchell, 2006) – all of which exhibit

responsiveness to stimuli and self-organization to maintain order whilst resource competing – the idea of organic visualisation becomes compelling for visual analysis of technical systems and processes which exhibit similar behavioural traits. Thus, in organic visualisation approaches the abstract entities employed for the information visualisation are allowed to develop, evolve, interact towards emergent behaviour, reproduce and eventually die out altogether, providing a visual support system for understanding and interpreting complexity (Ogawa & Ma, 2009). Initially, the organic information visualisation was developed as part of a rule based system, which produced a three dimensional and interactive network comprising unique words found in the text (Fry, 2000). Here, frequent words were used to create the surface of a 3D network envelope, whilst they were also grouped based on their location in the text. Another example of organic visualisation is the code-swarm environment (Ogawa & Ma, 2009), developed to qualitatively visualise an open-source project development process. Developers would commit their contributions to a central code repository, which would track their activity via logs making it possible to visualise which files were being committed, when and by whom. Outdated files would then fade away, emphasizing the end of a development cycle. In the context of this work, the dynamics of information seeking activity can be organically visualised with respect to the development, evolution and emergent spatial organisation, which is governed both by the inherited coding structure (see Table 1) and the force-field based layout.

4.1.1 Approach

If network growth is considered with respect to the addition of network elements (*i.e.* node and node's degree), then every activity within the recorded session can be divided into a finite number of steps i in which this addition occurred. For analysis purposes we consider both continuous and discrete network growth (Moody, et al., 2005), with the latter taking into account network configuration snapshots for each third of the total time period, showing the emerging interaction over the overall study.

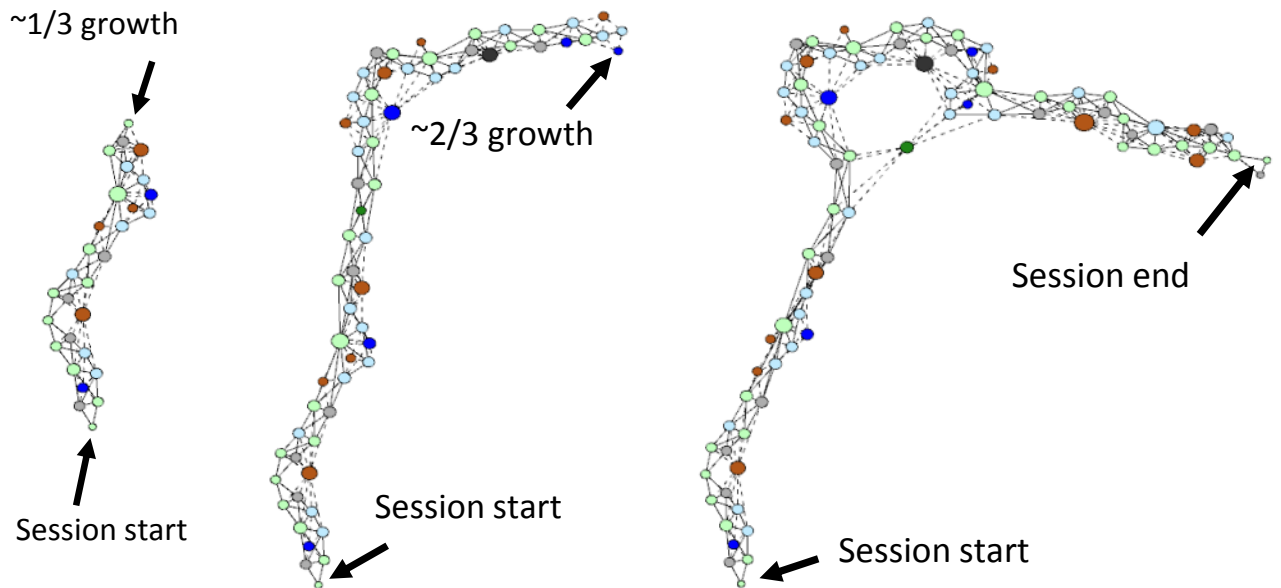


Figure 2: An example of network growth shown for each third of a session.

Thus, it is possible to visualise the dynamics of network growth by adding the corresponding elements at every step i . These are then made navigable within the visualisation environment. The nodes are either activities or sources while the relations are either temporal or plain mappings, as defined in Section 3. The outcome of such an approach is depicted in Figure 2, which shows the network's state for each third of a recorded session. The loop shown in the final third of the image indicates iteration carried out in the last portion of the session. The organic appearance of the process is due to the layout generation approach.

Based on the coding taxonomy (Table 1) and the principles used to establish the network (Section 2), the following patterns found in the example above (Figure 2) are also common to all the information seeking networks. These common features are shown in Figure 3 and will henceforth be referred to as motifs.

Motif **A** (Figure 3) depicts a regular chain of activities and information sources. Motif **B** denotes a local iteration, *i.e.* a node with a high degree of local connections, which in this case is a frequently accessed source. The other way in which a local hub is formed is to have a single operation that involves multiple sources. Further, motif **B** also illustrates revisiting a source in the duration of a single activity – forming multiple edges between ‘catalogue’ and activities such as ‘finding within source’. Finally, motif **C** exhibits a large-scale iteration involving a ‘wiki’ type source. The same motif also shows two local iterations involving ‘search engine’ and ‘technology article’ sources.

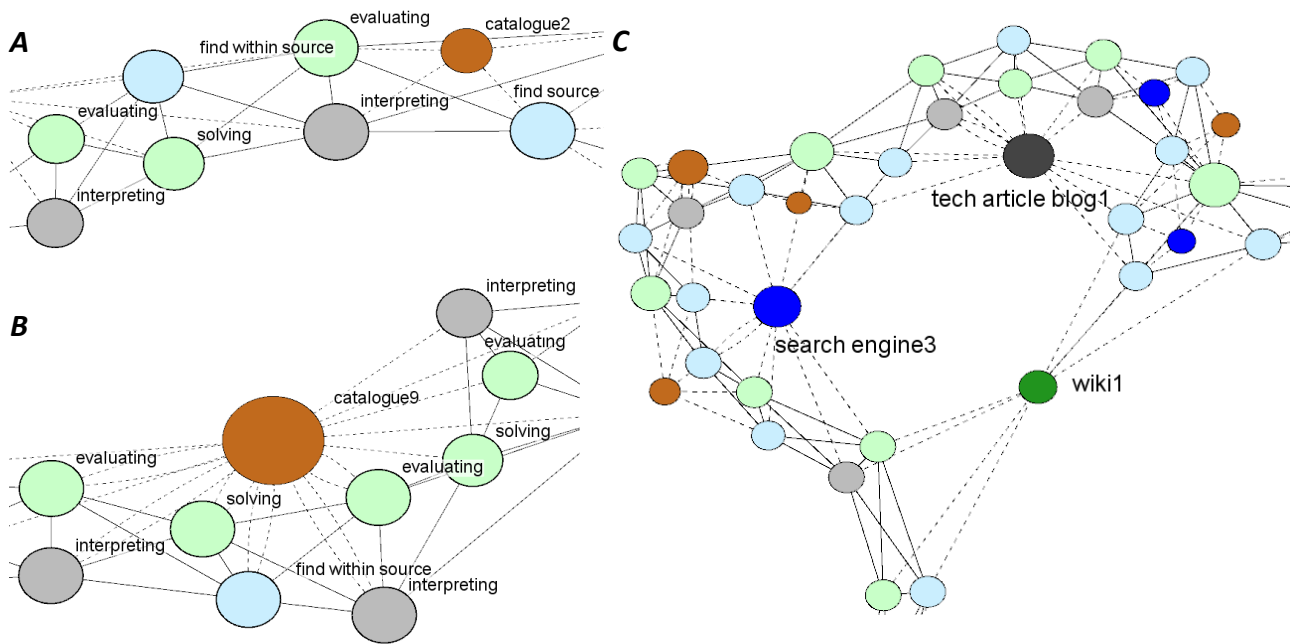
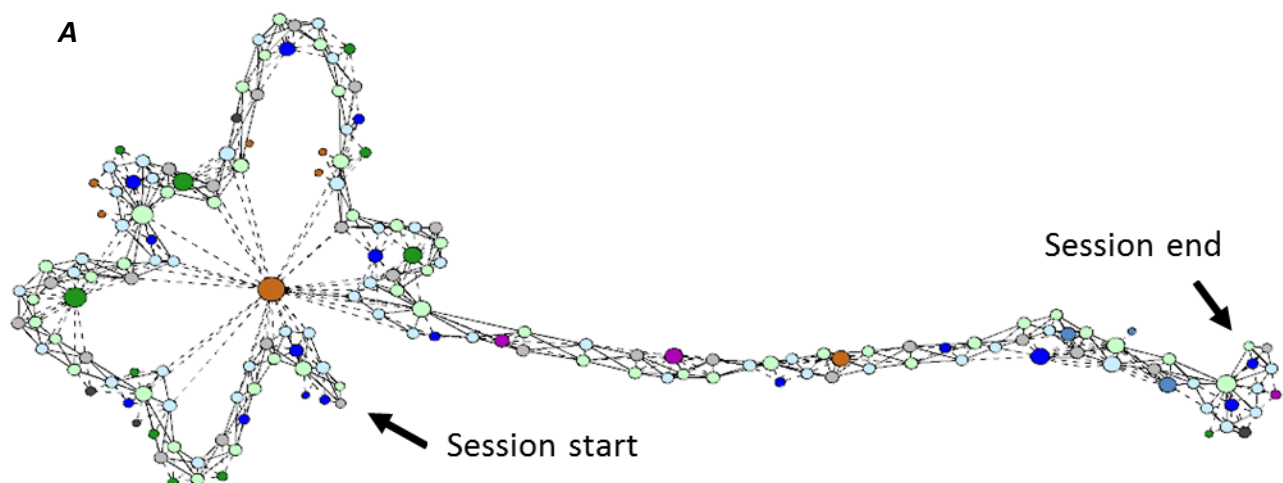
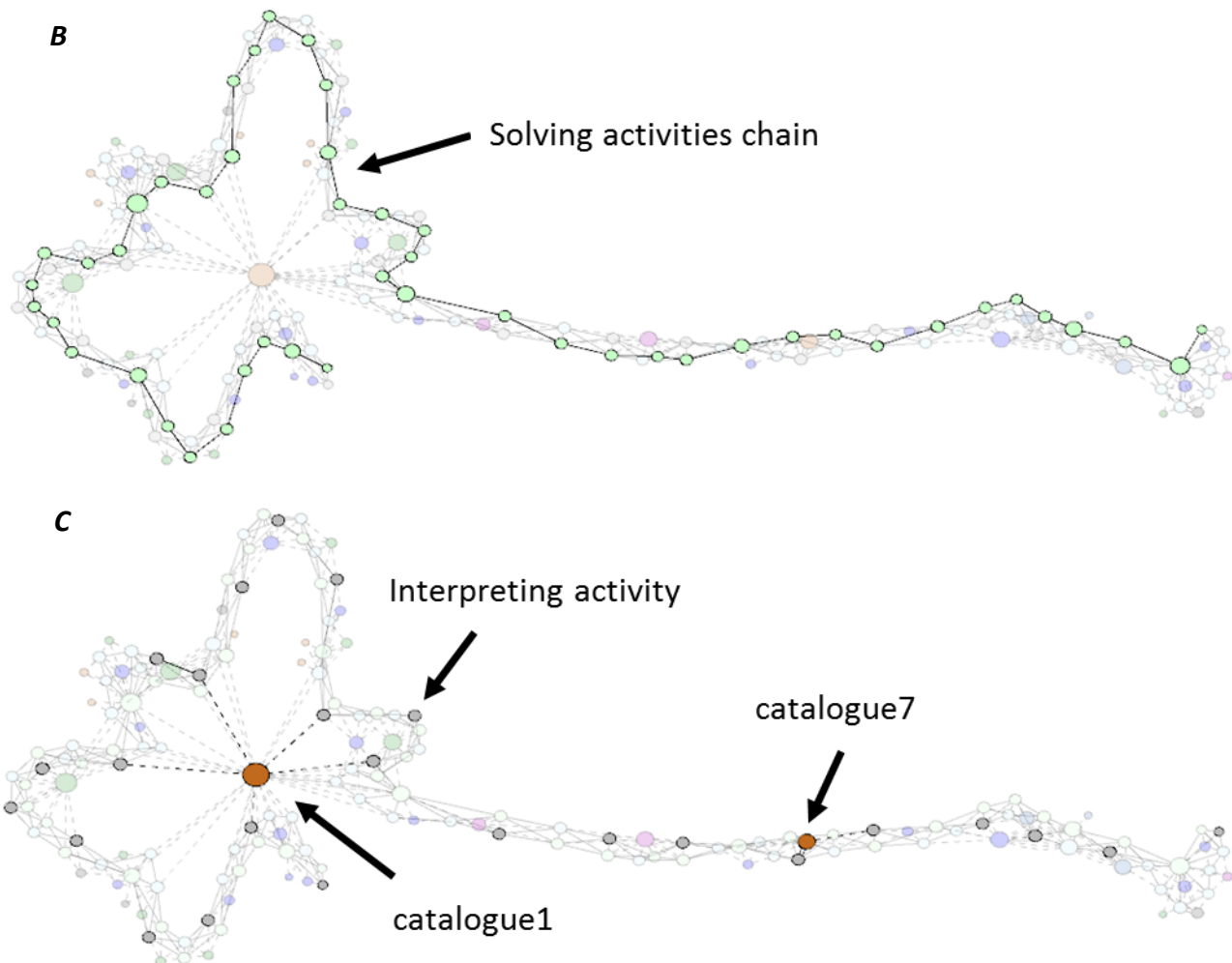


Figure 3: Three network motifs occurring in the information seeking networks.

4.1.2 Results

Figure 4 shows three perspectives on a network based on student participant 1 from the experimental study. Figure 4A shows the overall network, displaying the various activities and sources. From this it is possible to identify a number of features such as local iterations, regular chains and large-scale iterations. Figure 4B highlights the chain of solving activities, which can be used to guide the identification of periods of continuously linked activity. Finally, Figure 4C demonstrates how the network visualisation can be used to identify key sources. In this case selective filtering allows the researcher to examine only catalogues that are linked to interpretation activity.



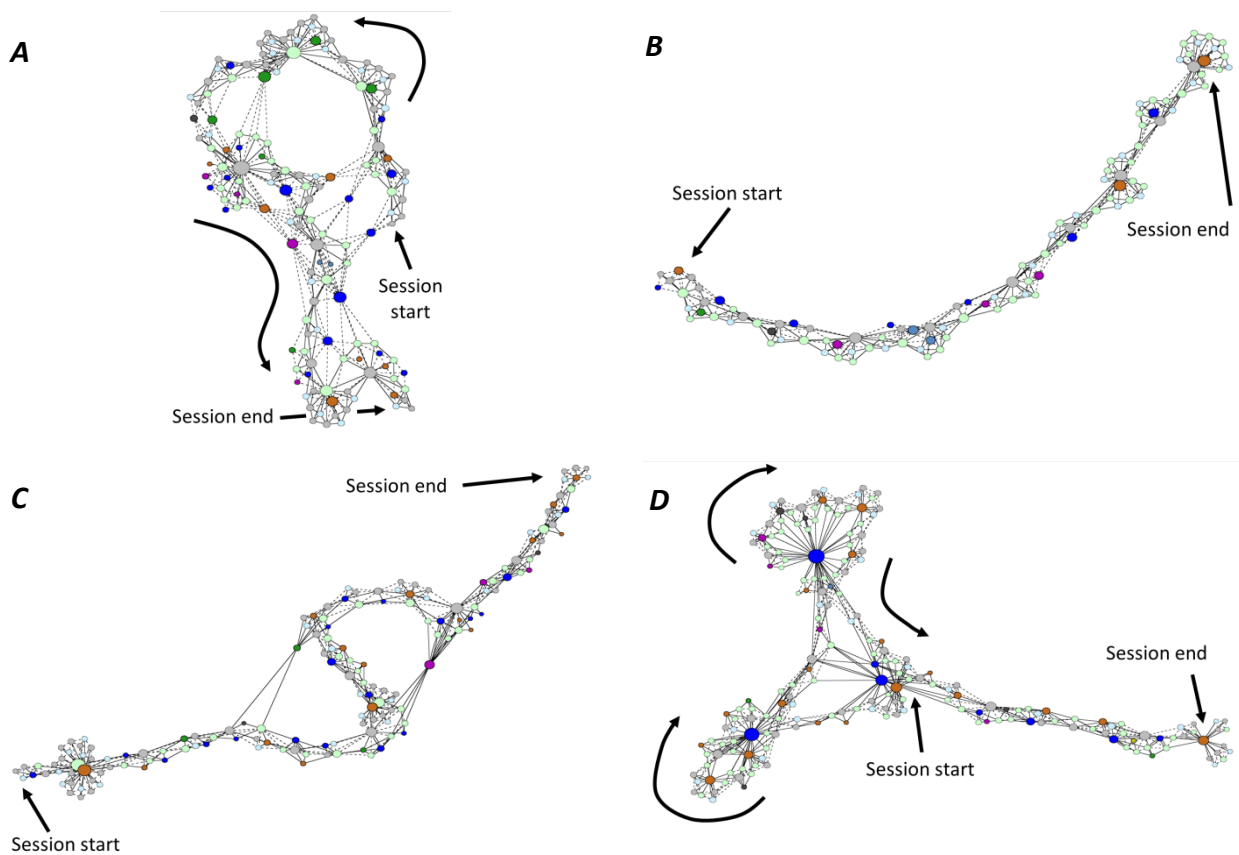


*Figure 4: Three perspectives on a network: **A** no filtering, **B** highlighting the solving chain and **C** highlighting only catalogues which were involved in interpreting activity.*

Bringing the different perspectives displayed in Figure 4 together allows for a detailed qualitative analysis of the designers activity. Specifically, it is clear that the large-scale iteration in the first half revolved round extensive catalogue use, while other sources were only locally referred to. The second half of the session then proceeds with only local iterations.

Finally, Figure 5 shows the four common patterns of activity observed during the studies. Using this representation allows activity patterns to be immediately and directly compared. This allows the identification of common patterns or differences that are not immediately obvious using single activity analysis alone. In particular, by identifying these four different patterns of information seeking activity it was possible to group participants and focus analytical effort on determining the reasons for the differences, which are not apparent when simply considering total duration of each activity/source or how these changed over time individually. More specifically, grouping was

achieved by counting the number of large-scale iterations in any given section of the network compared to the amount of nodes conforming to the regular chain motif (Figure 3). This in turn allowed analytical effort to be focused on understanding why these participants showed distinctly different network configurations and what these differences might mean. This is in comparison to the standard analysis of the number of sources or overall distribution of activities over time which showed no substantial differences, making further focused analysis difficult (Cash, et al., 2013). As such, this approach allows for the combination and collective consideration of multiple facets of design activity.



*Figure 5: Four typical patterns of activity: overall iterative **A**, local iteration only **B**, mixed **C** and a distinct separation between local iteration and overall iterative working phases **D**.*

4.2 Rate of Growth Analysis

Visualising the network using graphs conveys qualitatively the dynamic structure of the information seeking activity, allowing the analyst to focus on network motifs, activity patterns or network hubs – the most visited information sources. However, what remains hidden and undetectable for the analyst is the network dynamics, i.e. the temporal dependency of the emerging network's topology. Originally, the rate of change or pace of a dynamic network

(Moody, et al., 2005) was defined by edges (relations) with respect to levels (fast, slow), change (accelerating, decelerating), or stability (cascades, jumps and starts, etc.). In our case, the design activity is modelled using an evolving network with a continuously increasing number of edges and nodes. New activities, such as accessing a new information source or reusing an existing one, are events defined by the temporal interactions between the designer and the information sources, all of which consequently shape the network topology. Therefore, in order to communicate the dynamics of the information seeking activity we consider both edges and nodes in the rate of growth analysis. The intention of this work is to allow for the identification and quantitative characterisation of trends in activity. This is essential for performing comparisons or aggregating many networks and is considered a prerequisite for the probabilistic quantification of design activity patterns. Thus this analysis shifts away from statistical quantities describing the system e.g. mean degree, geodesic path length, clustering coefficients (Newman, 2003), random or law bound network evolution (Dorogovtsev & Mendes, 2002) since none of these address the micro level temporal interactions between designer and information sources (or other activities) which are the fundamental features of the design process. For the purpose of this example, the analysis is visually represented as a standard mathematical 2-D chart (of moving average established trends).

4.2.1 Approach

To address the problem of quantifying the dynamics of the information seeking activity, the approach described here focuses on observing the rate of growth of the network's edges with respect to the network's nodes over the study time steps. As such, two measures are proposed: the actual growth rate δ and the averaged and normalized network growth rate δ_e , which are defined in the following text.

For each time step i in which a node was added or a nodal degree has increased, with respect to the corresponding total number of edges m or total number of nodes, the actual growth δ is measured with the following expression:

$$\delta(i) = m(i) - n(i) \tag{1}$$

The measure δ_e takes into account the total size of the network at the end of the study, which is then averaged over all steps i . This is performed with respect to the total number of steps $i = p$ constituting the session as well as the total number of edges m_p and nodes n_p . Thus, for each step i the δ_e is defined by the following expression:

$$\delta_e(i) = i \cdot (m_p - n_p) / p \quad (2)$$

In this case, the majority of the studies showed an excess of edges over nodes, justifying the subtraction of nodes from edges rather than the other way around in order to have the bulk of values located above the abscise. Finally, the relative network growth indicator δ_q per step i is defined with respect to measures as given by Eq. (1) and (2):

$$\delta_q(i) = \delta_e(i) - \delta(i) \quad (3)$$

4.2.2 Results

Figure 6 shows a chart of relative network growth using data from practitioner participant 1 from the experimental study. The faint zigzag line corresponds to the relative network growth while the labelled thick line shows the moving average calculated for each 10 steps i established over 100% of recording session time.

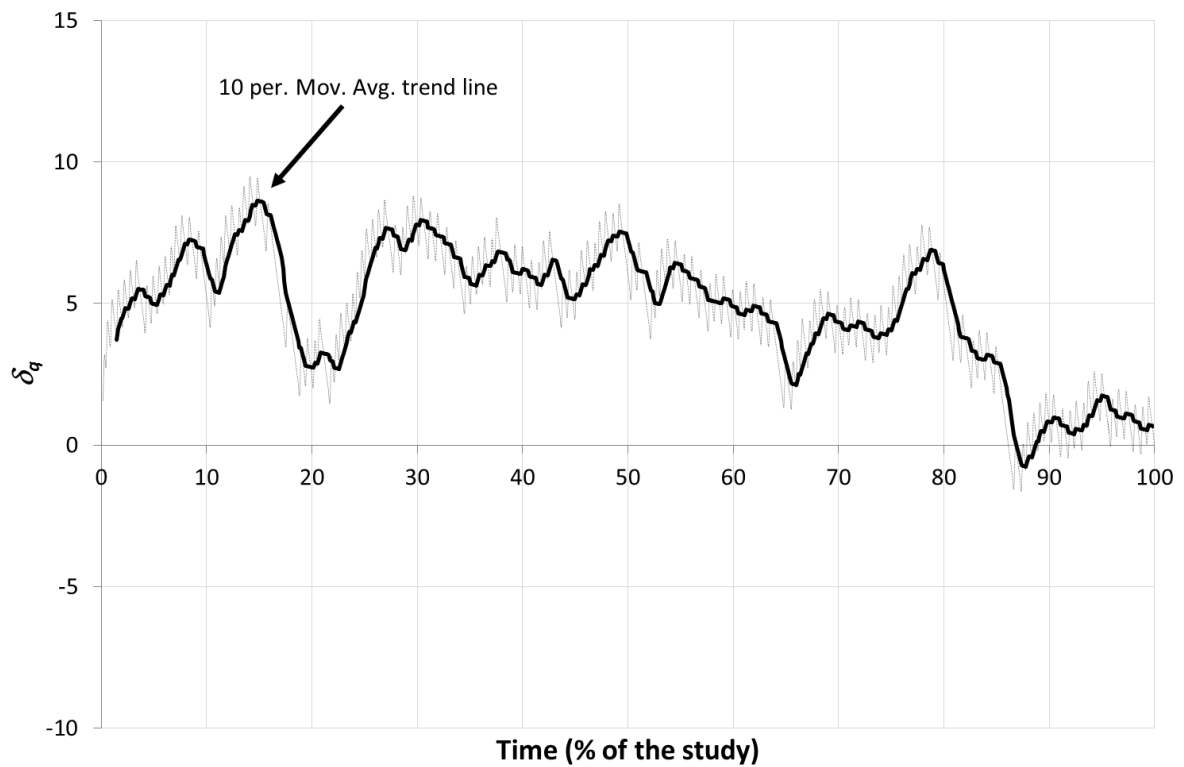


Figure 6: The relative network growth trend line established for practitioner participant 1 from the experimental study.

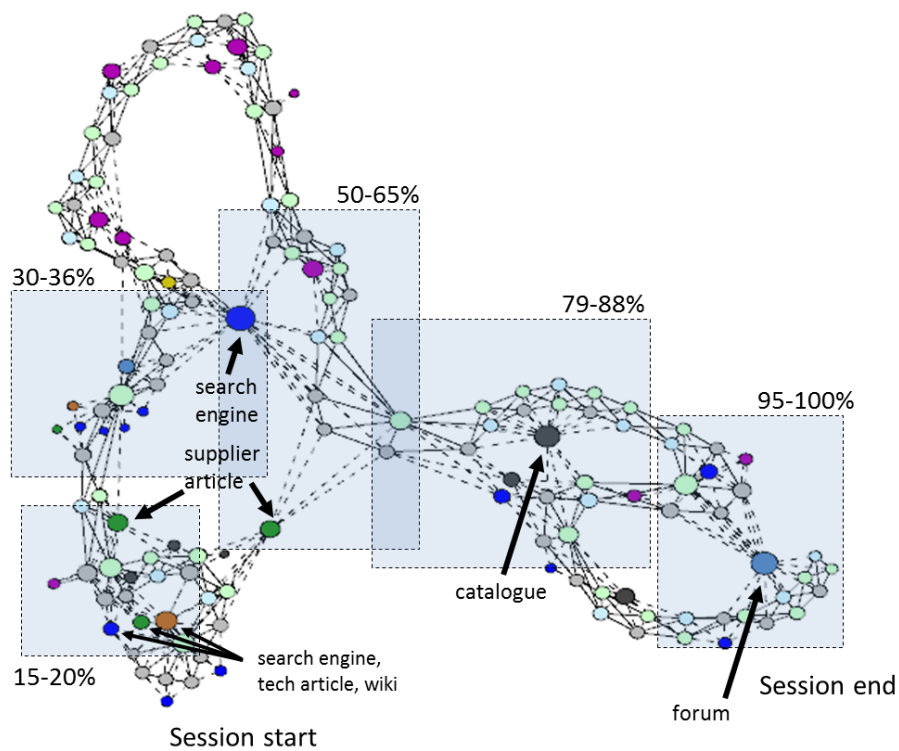


Figure 7: Correlation of the network structure to the rate of growth analysis shown in Figure 6 for practitioner participant 1 from the experimental study.

The positive trends in Figure 6 correspond to chains of activities, while the negative ones correspond to iteration related to revisiting the same information source or accessing multiple sources within the same activity. For instance, the negative trends in the relative growth correspond to the formation of the network motifs (Figure 3B and C) that indicate clustering with multiple sources revisited (between 15-20%) or iterations (between 79-88%), as shown in Figure 7. Further, by combining the two figures we can see that between 30-36% and 50-65% the same search engine source was used. As such, the combination of the network visualisation and 2nd order analysis provides a powerful tool for exploring the activity of designers.

Figure 8 shows how the rate of growth analysis can be used for comparison purposes – allowing quantitative comparisons to be made between the activities of different participants (in this case the three practitioners from the experimental study). It is also worth noting that by analysing activity in this way it is possible to build up statistically significant trends, which can be used to define patterns of interaction beyond the scope of traditional techniques, which tend to focus on individual activities rather than their interrelation. For example, the results highlighted here show a common downward trend with periods of similar and dissimilar network growth. Given sufficient data a normative model could be established and used to guide the use of training, stimuli or other techniques in order to disrupt the slowdown in network growth and to establish the features of successful versus unsuccessful periods of activity. In this sense Figure 9 has been used to illustrate what such a normative plot might look like and how it could be formulated such that comparisons can be made between individuals and the norm as well as against the ultimate outcome of the session. The example demonstrated in Figure 9 has been formed using the results for the twelve student participants, which were aggregated to give a basic trend with a probabilistic distribution. This demonstrates the potential of the method in this context and is not intended to be used as a normative model, as any such attempt would require significantly more participants.

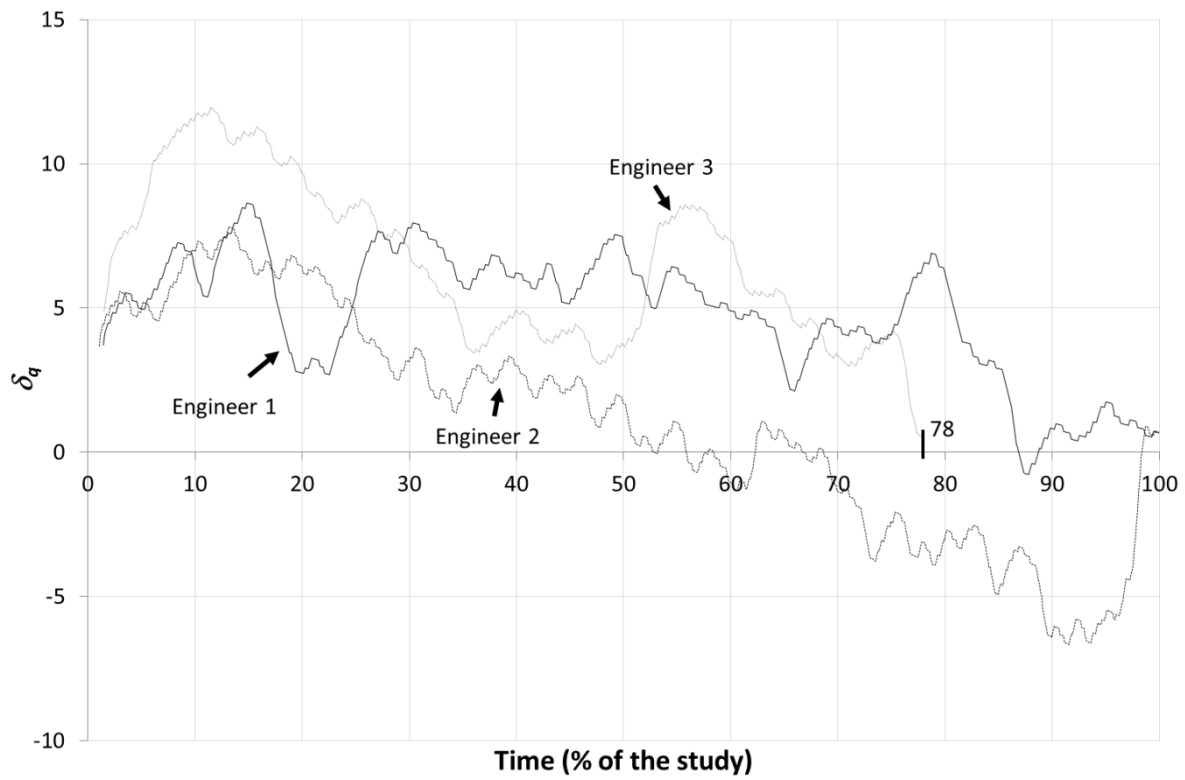


Figure 8: Comparison of trend lines for the three practitioner participants from the experimental study.

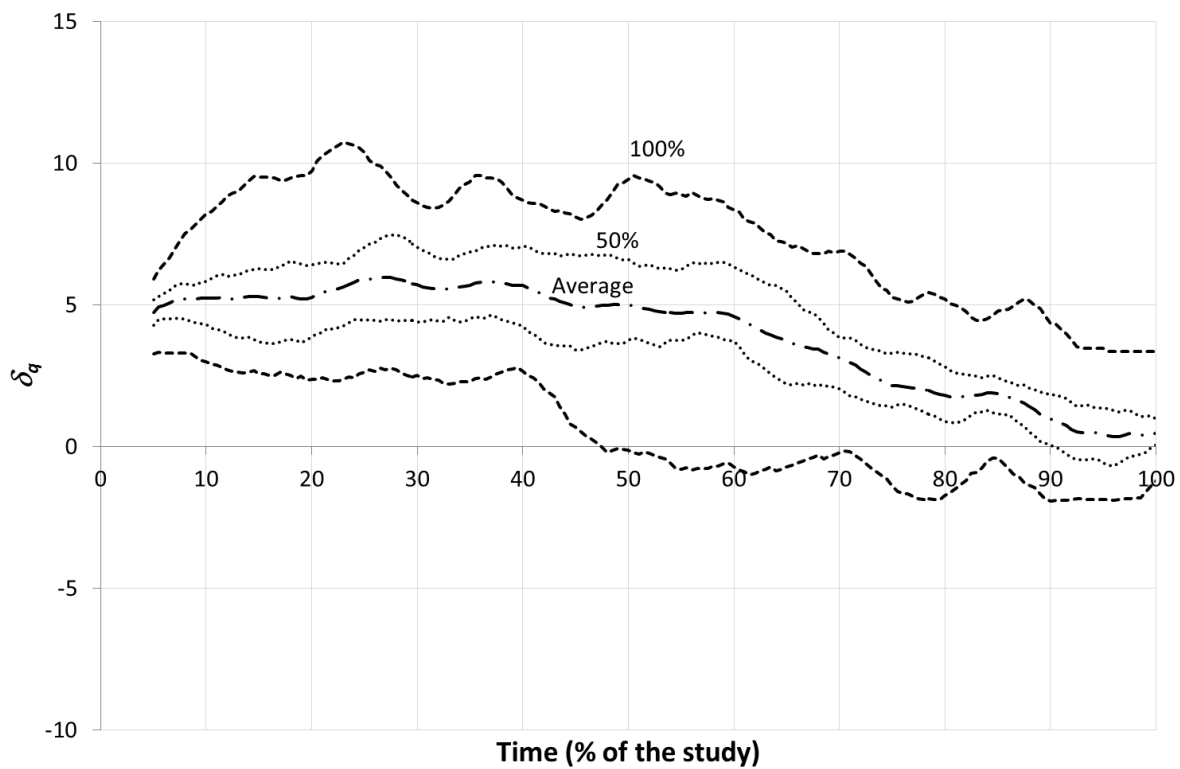


Figure 9: Results spread for twelve student participants from the experimental study.

5 Discussion

Current research on the analysis of protocol data related to designer activity typically relies on one of two approaches. The first is to consider each coded activity individually (*e.g.* Adams et al. (2011)), while the second is to link them to a wider theoretical framework – see for example Tang et al.'s (2011) use of the function-behaviour-structure framework in this context. In each of these cases there are clear difficulties in linking multiple variables and identifying patterns beyond those at the individual variable level. As such, it can be reasoned that although these approaches offer great potential for exploring the design situation they are limited in scope when used to identify relationships between activities, model complex networks of interrelated activities over time or provide deeper analysis of patterns of activity. In this context, it is possible to envisage using the proposed method to identify key motifs of design activity and subsequently characterise the 'health' of a design team, long-term design project or other such activity by monitoring the evolution of the network and relating key features to relevant external success criteria. In many ways this can be conceptually linked to the idea of characterising patterns of successful communication, where healthy activity can be differentiated from unsuccessful activity, predictions of future success made and overall motifs associated with various criteria (Maznevski & Chudoba, 2000).

In response to this need, this paper developed a method for using visual information analysis to explore complex patterns in the activity of designers. The method provides a means for quantitatively and qualitatively exploring the relationships between complex networks of activities and other variables. Further, the method was subsequently demonstrated on both experimental and observational data of designer's information seeking activity. In addressing this issue this method both supports existing analytical approaches – giving researchers another tool for exploring simple relationships between variables – and also opens up the possibility of more advanced rate of growth analysis and probabilistic prediction of complex design activity.

Finally, it is important to consider the generalisability of the method with respect to both designer activity and existing approaches to its analysis. In this regard the function-behaviour-structure (FBS) framework provides an initial example for how the method can be adapted to different approaches. In the general sense (Gero & Mc Neill, 1998) nodes could be characterised as either F, B or S while the edges or links between them could be defined as the various possible

transformations such as $F > F$ or $F > S$. Here, instead of using the posteriori temporal edges as in the information seeking example, edges are defined based on theory – one option highlighted in Step 2 of the proposed method. In a more specific example Kan et al. (2011) expands the FBS framework to include requirements, expected behaviour, behaviour from structure and description of the design. Here again, these could be described as nodes while the specific linkages considered by Kan et al. (e.g. $F > \text{expected behaviour}$) could be used to define edges, which could be in addition to temporal or other mappings. Again, the flexibility of the proposed method allows for this different perspective on designer activity. More generally, the method can be assessed in relation to the more fundamental characterisation of activity proposed in Activity Theory (Bedny & Harris, 2005). Here activity is decomposed over time by sequential layers of increasingly specific segments ending with unconscious operations and micro blocks. Using this general framework for the characterisation of activity – which is also consistent with both ‘move’ based (Goldschmidt & Tatts, 2005) and ‘episodic’ (Gero & Mc Neill, 1998) perspectives – gives a general means for assessing the proposed method. In this context it can be seen that nodes can be defined at each level of activity theory – as these are always segmented over time – whilst edges can be used to link these segments either based on direct physical/temporal mappings or on theoretical mappings. Although, this does not constitute a full validation of the methods efficacy for all possible aspects of, or perspectives on, designer activity, it strongly suggests that where activity can be segmented over time or an approach can be linked to Activity Theory the proposed method is likely to be applicable.

5.1 Contributions and Implications

There are three main implications from this work. Firstly, the method complements existing analysis methods in design research by providing rich qualitative information with respect to patterns of activity in the form of the network visualisations. Based on the examples given in Section 4, it is possible to see how this method could be used to help explore less linear models of activity where iterations or other such patterns are more dominant. Further, by providing the means to carry out second-order analysis of these patterns of activity, this method provides, for the first time in design research, a means for probabilistically modelling patterns of activity relating to multiple interrelated variables as opposed to attempting to combine separate models of each viable.

In addition to these contributions the method opens the door to a range of new research possibilities in the exploration and refinement of understanding with respect to existing work as well as future studies. In this context, by building on the standard model of protocol analysis, the method can be applied to the ever-increasing back catalogue of protocol studies – thus this offers a significant opportunity to help bring together existing datasets.

5.2 Limitations

Despite the strengths of the proposed method there are a number of limitations that should be considered. Primarily, the derived rate of growth data is a-contextual, showing only the magnitude of change in the network. As such, there are difficulties associated with establishing what exactly this data means with respect to a given dataset and how this might affect attempts to use this data to change practice. The rate of growth analysis method may also be limited to systems of a certain complexity or number of relationships before it becomes impossible to deconstruct – although this has not been established based on the work presented in this paper and therefore constitutes a key area of further research.

A second area that demands further work, but is substantial beyond the scope of this paper, is the full exhaustive validation of the method and its limitations in the context of design research foci, input dataset parameters (primarily complexity) and the limitations of the analysis techniques. Although the demonstration of the method based on the two studies outlined in Section 3 validates the methods applicability in this context, it cannot be considered a full validation and should instead be seen as an illustrative example.

Finally, the proposed method's generalisability across all types of design activity requires full validation based on numerous datasets. Ideally this could be achieved by examining a range of standard datasets, which have also been examined using other current approaches. However, the only resource that would approach this status in the design domain would be Delft protocol studies (Cross, et al., 1996) which in itself presents a limited body of data for analysis of this type. As such, it is envisioned that further validation work will be undertaken both by the authors and by the community in adoption of this method and the exploration of its application in different contexts. Indeed expanding the analysis to new design situations and coding approaches presents a significant body of work in its own right and is significantly beyond the scope of this paper, which

focuses on the proposition, demonstration and theoretical discussion of the visualisation method for design activity.

6 Conclusions

This paper proposes a method for using visual information analysis to explore complex patterns of activity in order to further the development of analysis techniques for protocol data (see Figure 1 for overall approach) – one of the most popular techniques used in design research today (Chai & Xiao, 2011). The core contribution of the proposed method is in supporting the analysis of patterns of interlinked variables over time rather than on an individual basis – which has already been noted as a strong practical need in design research.

The proposed method has several key contributions to field: it allows for detailed visualisation of multiple activities and other variables for a wide range of research foci. It supports the development of detailed probabilistic models of these interlinked activities by allowing for rate of growth analysis of the activity network. The method was designed to build on standard protocol analysis data meaning that it can be retrospectively applied to existing datasets in addition to providing a platform for future research. Finally, the benefits of the method for supporting cognitive offload when qualitatively analysing data (by allowing interdependencies to be readily identified across multiple variables and with respect to time) and providing more fundamental insight into the quantitative analysis of complex networks of activity (by allowing network growth to be explored over time) have been demonstrated, although full validation across all aspects of design activity presents significant future work.

In the context of the evolution of multifaceted networks of activity, the method helps to embed the complex traits of the systems that propel the evolution itself, in order to provide a visualisation, which truly augments and fosters the analysis process.

Acknowledgements

The work reported in this paper has been undertaken as part of the EPSRC Innovative Manufacturing Research Centre at the University of Bath (grant reference GR/R67507/0) in collaboration with the University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture (research project no. 120-1201829-1828 Models and Methods of Knowledge Management in Product Development). The authors gratefully acknowledge and express their thanks for the support of Prof. Ben Hicks and Prof. Steve Culley for financially supporting this

collaboration. Finally, the authors would like to express their thanks to the engaged and extremely useful comments of the reviewers.

References

- Adams, R. S., Daly, S. R., Mann, L. M., & Dall'Alba, G. (2011). Being a professional: Three lenses into design thinking, acting, and being. *Design Studies*, 32, 255-291.
- Allard, S., Levine, K. J., & Tenopir, C. (2009). Design engineers and technical professionals at work: Observing information usage in the workplace. *Journal of the American Society for Information Science and Technology*, 60, 443-454.
- Auriscchio, M., Bracewell, R., & Wallace, K. (2010). Understanding how the information requests of aerospace engineering designers influence information-seeking behaviour. *Journal of Engineering Design*, 21, 707-730.
- Barnes, J., & Hut, P. (1986). A hierarchical $O(N \log N)$ force-calculation algorithm. *Nature*, 324.
- Bedny, G. Z., & Harris, S. R. (2005). The systemic-structural theory of activity: Applications to the study of human work. *Mind, culture, and Activity*, 12, 128-147.
- Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of documentation*, 38, 61-71.
- Berland, G. K., Elliott, M. N., Morales, L. S., Algazy, J. I., Kravitz, R. L., Broder, M. S., Kanouse, D. E., Munoz, J. A., Puyol, J.-A., & Lara, M. (2001). Health information on the Internet. *JAMA: the journal of the American Medical Association*, 285, 2612-2621.
- Blandford, A., & Attfield, S. (2010). Interacting with information. *Synthesis Lectures on Human-Centered Informatics*, 3, 1-99.
- Borlund, P. (2003). The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information research*, 8, 8-11.
- Cash, P., Hicks, B. J., & Culley, S. (2013). A comparison of designer activity using core design situations in the laboratory and practice. *Design Studies*, <http://dx.doi.org/10.1016/j.destud.2013.03.002>.
- Cave, P. R., & Noble, C. E. I. (1986). Engineering design data management. In *1st International Conference on Engineering Management, Theory and Applications*. Swansea, UK.
- Chai, K. H., & Xiao, X. (2011). Understanding design research: A bibliometric analysis of Design Studies (1996-2010). *Design Studies*, 33, 24-43.
- Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical review E*, 70, 066111.
- Coley, F., Houseman, O., & Roy, R. (2007). An introduction to capturing and understanding the cognitive behaviour of design engineers. *Journal of Engineering Design*, 18, 311-325.
- Court, A. W., Ullman, D. G., & Culley, S. J. (1998). A comparison between the provision of information to engineering designers in the UK and the USA. *International Journal of Information Management*, 18, 409-425.
- Cross, N., Christiaans, H., & Dorst, K. (1996). *Analysing design activity*. Chichester: John Wiley and Sons, UK.
- Dong, A., Kleinsmann, M. S., & Deken, F. (2012). Investigating design cognition in the construction and enactment of team mental models. *Design Studies*, 34, 1-33.
- Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in physics*, 51, 1079-1187.
- Dorst, K., & Dijkhuis, J. (1995). Comparing paradigms for describing design activity. *Design Studies*, 16, 261-274.
- Fry, B. J. (2000). *Organic information design*. Massachusetts Institute of Technology.
- Gero, J. S., & Mc Neill, T. (1998). An approach to the analysis of design protocols. *Design Studies*, 19, 21-61.

- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*: Aldine de Gruyter.
- Goldschmidt, G. (1992). Criteria for design evaluation: A process-oriented paradigm. In Y. Kalay (Ed.), *Principals of Computer-Aided Design: Evaluating and predicting design performance* (pp. 67-79). New York: John Wiley & Sons Inc.
- Goldschmidt, G., & Tatsa, D. (2005). How good are good ideas? Correlates of design creativity. *Design Studies*, 26, 593-611.
- Goodman-Deane, J., Langdon, P., & Clarkson, J. (2010). Key influences on the user-centred design process. *Journal of Engineering Design*, 21, 345-373.
- Hagedorn, J., Hailpern, J., & Karahalios, K. G. (2008). VCode and VData: illustrating a new framework for supporting the video annotation workflow. In *Conference on Advanced Visual Interfaces* (pp. 317-321). New York, USA: ACM.
- Harary, F., & Gupta, G. (1997). Dynamic graph models. *Mathematical and Computer Modelling*, 25, 79-87.
- Hertzum, M., & Pejtersen, A. M. (2000). The information-seeking practices of engineers: searching for documents as well as for people. *Information Processing & Management*, 36, 761-778.
- Holscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer networks*, 33, 337-346.
- Huet, G., Culley, S. J., McMahon, C., & Fortin, C. (2007). Making sense of engineering design review activities. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing : AI EDAM*, 21, 243-266.
- Hyldegard, J. (2009). Beyond the search process: Exploring group members' information behaviour in context. *Information Processing & Management*, 45, 142-158.
- Jadad, A. R., & Gagliardi, A. (1998). Rating health information on the Internet. *JAMA: the journal of the American Medical Association*, 279, 611-614.
- Kan, J. W. T., & Gero, J. S. (2008). Acquiring information from linkography in protocol studies of designing. *Design Studies*, 29, 315-337.
- Kan, J. W. T., Gero, J. S., & Tang, H.-H. (2011). Measuring cognitive design activity changes during an industry team brainstorming session. In J. S. Gero (Ed.), *Design Computing and Cognition '10* (pp. 621-640): Springer.
- Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterising Web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology*, 58, 999-1018.
- King, D. W., Casto, J., & Jones, H. (1994). *Communication by engineers: a literature review of engineers' information needs, seeking processes, and use*. Washington, DC: Council on Library Resources
- Kwasitsu, L. (2004). Information-seeking behavior of design, process, and manufacturing engineers. *Library & Information Science Research*, 25, 459-476.
- Lethbridge, T. C., Sim, S. E., & Singer, J. (2005). Studying software engineers: Data collection techniques for software field studies. *Empirical software engineering*, 10, 311-341.
- LiveScribe. (2011). LiveScribe: never miss a word. In: <http://www.livescribe.com/en-us/>.
- Maznevski, M. L., & Chudoba, K. M. (2000). Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science*, 11, 473-492.
- Mitchell, M. (2006). Complex systems: Network thinking. *Artificial Intelligence*, 170, 1194-1212.
- Moody, J., McFarland, D., & Bender-deMoll, S. (2005). Dynamic network visualization. *American Journal of Sociology*, 110, 1206-1241.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM review*, 45, 167-256.

- Nguyen, P. T., Steinert, M., Carroll, A., & Leifer, L. (2012). Applying bioinformatics analysis principles to CAD data to better characterize and improve the engineering design process. In *Design 2012* (pp. 1049-1060). Dubrovnik, Croatia.
- Ogawa, M., & Ma, K. L. (2009). Code_swarm: a design study in organic software visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 15, 1097-1104.
- Oh, S., Oh, J. S., & Shah, C. (2009). The use of information sources by Internet users in answering questions. *Proceedings of the American Society for Information Science and Technology*, 45, 1-13.
- Panopto. (2011). Panopto - Capture.Publish.Share. In: <http://www.panopto.com/>.
- Parmee, I. C., & Abraham, J. A. (2004). User-centric evolutionary design. In *Design 2004* (pp. 1441-1446). Dubrovnik, Croatia.
- Pavkovic, N., Storga, M., Bojcetic, N., & Marjanovic, D. (2013). Facilitating design communication through engineering information traceability. *AI EDAM (Artificial Intelligence for Engineering Design, Analysis and Manufacturing)*, 27, 91-105.
- Pedgley, O. (2007). Capturing and analysing own design activity. *Design Studies*, 28, 463-483.
- Peterson, R. A., & Merino, M. C. (2003). Consumer information search behavior and the Internet. *Psychology and Marketing*, 20, 99-121.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioural research: A critical review of the literature and recommended remedies. *Journal of applied psychology*, 88, 879-903.
- Purchase, H., Andrienko, N., Jankun-Kelly, T., & Ward, M. (2008). Theoretical foundations of information visualization. *Information Visualization*, 4950, 46-64.
- Puttre, M. (1991). Product data management. *Mechanical Engineering*, 113, 81-83.
- Reed, N., Scanlan, J., Wills, G., & Halliday, S. T. (2011). Knowledge use in an advanced manufacturing environment. *Design Studies*, 32, 292-312.
- Robinson, M. A. (2010). An empirical analysis of engineers' information behaviours. *Journal of the American Society for Information Science and Technology*, 61, 640-658.
- Salustri, F. A., Eng, N. L., & Weerasinghe, J. S. (2008). Visualizing information in the early stages of engineering design. *Computer-Aided Design & Applications*, 5, 697-714.
- Stankovic, T., Storga, M., Stojic, I., & Savsek, T. (2012). Traceability visualisation toolkit. In *Design 2012*. Dubrovnik, Croatia.
- Storga, M., Marjanovic, D., & Savsek, T. (2011). Reference model for traceability records implementation in engineering design environment. In *ICED 11 International conference on engineering design*. Copenhagen, Denmark.
- Storga, M., Stankovic, T., Cash, P., & McAloone, T. (2013). Visually Augmented Analysis of Socio-Technical Networks in Engineering Systems Design Research. In *The Future of Transdisciplinary Design*. Luxembourg city, Luxembourg.
- Tang, H. H., Lee, Y. Y., & Gero, J. S. (2011). Comparing collaborative co-located and distributed design processes in digital and traditional sketching environments: A protocol study using the function-behaviour-structure coding scheme. *Design Studies*, 32, 1-29.
- Torgerson, D. J., & Torgerson, C. J. (2003). Avoiding bias in randomised controlled trials in educational research. *British journal of educational studies*, 51, 36-45.
- Unwin, A., Theus, M., & Hofmann, H. (2006). *Graphics of large datasets: visualizing a million*. New York, USA: Springer-Verlag.
- Vande Moere, A., Dong, A., & Clayden, J. (2008). Visualising the social dynamics of team collaboration. *Co-Design*, 4, 151-171.
- VCode. (2011). VCode and VData homepage [Online]. In. Illinois: <http://social.cs.uiuc.edu/projects/vcode.html>.

- Waddell, P. (2002). Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association*, 68, 297-314.
- Ware, C. (2004). *Information visualization: perception for design* (2nd ed. Vol. 22). San Francisco, USA: Morgan Kaufmann.
- Wasiak, J., Hicks, B. J., Newnes, L., Dong, A., & Burrow, L. (2010). Understanding engineering email: the development of a taxonomy for identifying and classifying engineering work. *Research in Engineering Design*, 21, 43-64.
- White, S. (2011). Business population estimates for the UK and regions 2011. In B. a. I. Skills (Ed.). Sheffield.